CLAIMS

WE CLAIM:

1. A method of updating an electronic document for information retrieval comprising the steps of:

(a) dividing the electronic document into a plurality of partitions;

(b) receiving for the electronic document, update portions;

5   (c) matching the update portions to at least one partition;

(d) concurrently locking at least one partition for updating while keeping at least one partition unlocked for reading;

(e) updating the locked partition with the update portions only if the keyword matches with the locked partition; and

10   (f) changing the locked and unlocked partitions and repeating steps (d) and (e) to update each of the partitions over a predetermined period;

whereby the electronic document may be updated concurrently with use of the electronic document.

2. The method of claim 1 wherein the electronic document is an index linking a set of keywords to document identifiers, and the update portions are keywords and document identifiers for a new document and wherein at step (c) it is the keyword that is matched to a partition.

3   The method of claim 2 further including the steps of:

(g) matching portions of received queries to at least one partition;

(h) reading partitions matched to the portions to respond to the query.

4. The method of claim 3 wherein the matching of the portions of the queries to at least one partition of step (f) and the matching of the keywords for the new document to at least one partition of step (c) use a common mapping means.

5. The method of claim 4 wherein the mapping means is a hash table accepting the keyword as an argument to produce a partition as a value.

6. The method of claim 2 further including the step of storing the keywords and document identifiers for a new document in a change-log file before step (e) of updating the locked partition with the keyword and document identifiers for the new document.

7. The method of claim 5 wherein the change-log file includes a time stamp indicating the time of storing the keywords and document identifiers for a new document and the partitions include a time stamp indicating when the partition was last updated and wherein the step of updating the partition reads entries of the change-log file having a time stamp later than the time stamp of the partition and then updates the time stamp of the partition;

    whereby updates are ensured to have been recorded in the change-log file.

8. The method of claim 2 wherein the partitions include an main portion stored in a first storage device having a first access speed and a supplemental portion stored in a second storage device having a second access speed faster than the first access speed, and wherein step (e) updates the supplemental portion of the locked partition, and further including the steps of:

    (g) causing queries of the index to read both the index portion and the supplemental portion; and

    (h) at predetermined intervals, merging the main portion with the supplemental portion.

9. The method of claim 8 predetermined period of time at which partitions are updated with the keywords and documents for the new document is less than the periodic interval when the main portion of the partition is updated with the supplemental portion of the partition.

10. The method of claim 8 wherein the main portions larger than the supplemental portion.

11. The method of claim 8 wherein the first storage device is a disk drive and the second storage device is solid-state memory.

12. The method of claim 8 wherein the predetermined interval is selected from the group consisting of: a periodic interval based on the amount of data stored in the supplemental portion, a constant periodic interval, and a periodic interval based on the partition.

13. The method of claim 8 wherein the merging of step (h) compacts the combined supplemental portion and main portion.

14. The method of claim 8 wherein the merging of step (h) computes global statistics of the combined supplemental portion and main portion.

15. The method of claim 8 wherein the merging of step (h) includes the steps of:

(i) freezing the supplemental portion and designating a second supplemental portion for receiving new keywords and document identifiers for new documents;

(ii) combining the frozen supplemental portion and the main portion to create a second main portion; and

(iii) deleting the frozen supplemental portion and the main portion and using the second supplemental portion as the supplemental portion and using the second main portion as the main portion.

16. The method of claim 14 wherein during step (i) queries are directed to the frozen supplemental portion, the second supplemental portion and the main portion.

17. The method of claim 14 wherein the supplemental portion and the main portions are identified by pointers and wherein step (iii) of using the second supplemental portion as the supplemental portion and using the second main portion as the main portion is accomplished by redirecting pointers.

18. The method of claim 8 further including the step of storing the keywords and document identifiers for a new document in a change-log file before step (e) of updating the locked partition with the keyword and document identifiers for the new document.

19. The method of claim 18 wherein the change-log file includes a time stamp indicating the time of storing the keywords and document identifiers for a new document and the partitions portion include a time stamp indicating when the partition portion was last updated and wherein the step of updating the partition portion updates the time stamp

5   of the partition portion;

whereby loss of the supplemental portion may be remedied by reference to the change-log file.

20. The method of claim 18 wherein the step (e) of updating the locked partition with the keyword and document identifiers includes the step of reviewing each partition for the earliest time stamp and deleting from the change-log file all keywords having an earlier timestamp than the earliest time stamp for all partitions.

21. A method of updating an electronic document for information retrieval, the electronic document including a main portion stored in a first storage device having a given access speed, the method comprising the steps of:

(a) providing a supplemental portion in a second storage device having an access

5   speed faster than the first storage device;

(b) storing, updates of the electronic document in the supplemental portion;

(c) causing queries of the electronic document to read both the supplemental portion and the main portion; and

(d) at predetermined intervals, merging the main portion with the supplemental

10   portion;

whereby the electronic document may be updated concurrently with use.

22. The method of claim 21 wherein the electronic document is an index linking keywords to document identifiers and the update portion is linked keywords and document identifiers for a new document.

23. The method of claim 22 wherein the main portion is larger than the supplemental portion.

24. The method of claim 22 wherein the first storage device is a disk drive and the second storage device is solid-state memory.

25. The method of claim 22 wherein the predetermined interval is selected from the group consisting of: a periodic interval based on the amount of data stored in the supplemental portion and a constant periodic interval.

26. The method of claim 22 wherein the merging of step (d) compacts the combined supplemental portion and main portion.

27. The method of claim 22 wherein the merging of step (d) computes global statistics of the combined supplemental portion and main portion.

28. The method of claim 22 wherein the merging of step (d) includes the steps of:

(i) freezing the supplemental portion and designating a second supplemental portion for receiving new keywords and document identifiers for new documents;

(ii) combining the frozen supplemental portion and the main portion to create a second main portion;

(iii) deleting the frozen supplemental portion and the main portion and using the second supplemental portion as the supplemental portion and using the second main portion as the main portion.

29. The method of claim 28 wherein during step (i) queries are directed to the frozen supplemental portion, the second supplemental portion and the main portion.

30. The method of claim 28 wherein the supplemental portion and the main portions are identified by pointers and wherein step (iii) of using the second supplemental portion as the supplemental portion and using the second main portion as the main portion is accomplished by redirecting pointers.

31. The method of claim 22 further including the step of storing the keywords and document identifiers for a new document in a change-log file before step (d) of updating the main portion with the supplemental portion.

21

32. The method of claim 31 wherein the change-log file includes a time stamp indicating the time of storing the keywords and document identifiers for a new document and the main portion include a time stamp indicating when the main portion was last updated and wherein the step of updating the main portion updates the time stamp of the

5   main portion;

whereby loss of the supplemental portion may be remedied by reference to the change-log file.

33. The method of claim 31 wherein the step (d) of updating the main portion includes the step of deleting from the change-log file all keywords having an earlier timestamp than that of the main portion.

34. The method of claim 22 including the further steps of:

dividing the main and supplemental portions of the index into a plurality of partitions;

at step (b) storing keywords and document identifiers for the new document a

5   predetermined partition of the supplemental portion;

at the predetermined intervals, sequentially merging the partitions of the supplemental portion with corresponding partitions of the main portion;

receiving bulk-load keywords and document identifiers for the index;

pre-dividing the bulk-load keywords and document identifiers into partitioned

10   files related to the partitions of the main and supplemental portions of the index;

sequentially storing a partitioned file in the second storage device and merging the partition file with the corresponding partition of the main portion;

whereby bulk-load data may be efficiently integrated with the index.

35. The method of claim 22 wherein the partition file is merged with the corresponding partition of the main portion at a second predetermined interval different from the first predetermined interval.

36. A system for information retrieval comprising:

an electronically readable document divided into a plurality of partitions;

a program executed on an electronic computer and communicating with the electronically readable document to:

5   (a) receiving update portions;

(b) matching the update portions to at least one partition;

(c) concurrently lock at least one partition for updating while keeping at least one partition unlocked for reading;

(d) update the locked partition with the keyword and document identifiers for the

10   new document only if the keyword matches with the locked partition; and

(e) change the locked and unlocked partitions and repeat steps (d) and (e) to update each of the partitions over a predetermined period;

whereby the electronic document may be updated concurrently with use of the index.

37. The system of claim 36 wherein the electronic document is an index linking a set of keywords to document identifiers, and the update portions are keywords and document identifiers for a new document and wherein the program matches an update portion to a partition through the keyword.

38 The system of claim 37 wherein the program further executes the steps of:

(f) matching portions of the received queries to at least one partition;

(g) reading partitions matched to the portions to respond to the query.

39. The system of claim 38 wherein the matching of the portions of the queries to at least one partition of step (f) and the matching of matching of the keywords for the new document to at least one partition of step (c) use a common mapping means.

40. The system of claim 39 wherein the mapping means is a hash table accepting the keyword as an argument to produce a partition as a value.

41. The system of claim 37 further wherein the program further executes the step of storing the keywords and document identifiers for a new document in a change-log file before step (e) of updating the locked partition with the keyword and document identifiers for the new document.

42. The system of claim 41 wherein the change-log file includes a time stamp indicating the time of storing the keywords and document identifiers for a new document and the partitions include a time stamp indicating when the partition was last updated and wherein the step of updating the partition reads entries of the change-log file having a time stamp later than the time stamp of the partition and then updates the time stamp of the partition;

>whereby updates are ensured to have been recorded in the change-log file.

43. The system of claim 37 wherein the partitions include an main portion stored in a first storage device having a first access speed and a supplemental portion stored in a second storage device having a second access speed faster than the first access speed, and wherein the program executes step (e) to updates the supplemental portion of the locked partition, and further including the steps of:

>(g) causing queries of the index to read both the index portion and the supplemental portion; and

>(f) at predetermined intervals, merging the main portion with the supplemental portion.

44. The system of claim 43 predetermined period of time at which partitions are updated with the keywords and documents for the new document is less than the periodic interval when the main portion of the partition is updated with the supplemental portion of the partition.

45. The system of claim 43 wherein the main portions larger than the supplemental portion.

46. The system of claim 43 wherein the first storage device is a disk drive and the second storage device is solid-state memory.

47. The system of claim 43 wherein the predetermined interval is selected from the group consisting of: a periodic interval based on the amount of data stored in the supplemental portion, a constant periodic interval, and a periodic interval based on the partition.

48. The system of claim 43 wherein the merging of step (f) compacts the combined supplemental portion and main portion.

49. The system of claim 43 wherein the merging of step (f) computes global statistics of the combined supplemental portion and main portion.

50. The system of claim 43 wherein the merging of step (f) includes the steps of:

(i) freezing the supplemental portion and designating a second supplemental portion for receiving new keywords and document identifiers for new documents;

(ii) combining the frozen supplemental portion and the main portion to create a second main portion; and

(iii) deleting the frozen supplemental portion and the main portion and using the second supplemental portion as the supplemental portion and using the second main portion as the main portion.

51. The system of claim 50 wherein during step (i) queries are directed to the frozen supplemental portion, the second supplemental portion and the main portion.

52. The system of claim 50 wherein the supplemental portion and the main portions are identified by pointers and wherein step (iii) of using the second supplemental portion as the supplemental portion and using the second main portion as the main portion is accomplished by redirecting pointers.

53. The system of claim 43 further wherein the program further executes the step of storing the keywords and document identifiers for a new document in a change-log file before step (e) of updating the locked partition with the keyword and document identifiers for the new document.

54. The system of claim 53 wherein the change-log file includes a time stamp indicating the time of storing the keywords and document identifiers for a new document and the partitions portion include a time stamp indicating when the partition portion was last updated and wherein the step of updating the partition portion updates the time stamp of the partition portion;

whereby loss of the supplemental portion may be remedied by reference to the change-log file.

55. The system of claim 53 wherein the step (e) of updating the locked partition with the keyword and document identifiers includes the step of reviewing each partition for the earliest time stamp and deleting from the change-log file all keywords having an earlier timestamp than the earliest time stamp for all partitions.

56. An system allowing on-line updating and comprising:

an electronically readable document including a main portion stored in a first storage device having a given access speed and a supplemental portion in a second storage device having an access speed faster than the first storage device;

5   an electronic computer communicating with the electronically readable document and executing a stored program to:

(a) store update portions in the supplemental portion;

(b) cause queries of the electronic document to read both the main portion and the supplemental portion; and

10   (c) at predetermined intervals, merging the main portion with the supplemental portion;

whereby the electronic document may be updated concurrently with use.

57. The system of claim 56 wherein the electronic document is an index linking a set of keywords to document identifiers, and the update portions are keywords and document identifiers for a new document.

58. The system of claim 57 wherein the main portion is larger than the supplemental portion.

59. The system of claim 57 wherein the first storage device is a disk drive and the second storage device is solid-state memory.

60. The system of claim 57 wherein the predetermined interval is selected from the group consisting of: a periodic interval based on the amount of data stored in the supplemental portion and a constant periodic interval.

61. The system of claim 57 wherein the merging of step (c) compacts the combined supplemental portion and main portion.

62. The system of claim 57 wherein the merging of step (c) computes global statistics of the combined supplemental portion and main portion.

63. The system of claim 57 wherein the merging of step (c) includes the steps of:

(i) freezing the supplemental portion and designating a second supplemental portion for receiving new keywords and document identifiers for new documents;

(ii) combining the frozen supplemental portion and the main portion to create a second main portion;

(iii) deleting the frozen supplemental portion and the main portion and using the second supplemental portion as the supplemental portion and using the second main portion as the main portion.

64. The system of claim 63 wherein during step (i) queries are directed to the frozen supplemental portion, the second supplemental portion and the main portion.

65. The system of claim 63 wherein the supplemental portion and the main portions are identified by pointers and wherein step (iii) of using the second supplemental portion as the supplemental portion and using the second main portion as the main portion is accomplished by redirecting pointers.

66. The system of claim 57 further wherein the program further executes the step of storing the keywords and document identifiers for a new document in a change-log file before step (d) of updating the main portion with the supplemental portion.

68. The system of claim 66 wherein the change-log file includes a time stamp indicating the time of storing the keywords and document identifiers for a new document and the main portion include a time stamp indicating when the main portion was last updated and wherein the step of updating the main portion updates the time stamp of the main portion;

whereby loss of the supplemental portion may be remedied by reference to the change-log file.

68. The system of claim 66 wherein the step (c) of merging the main portion and supplemental portions includes the step of deleting from the change-log file all keywords having an earlier timestamp than that of the main portion.

69. The system of claim 57 including the further steps of:

dividing the main and supplemental portions of the index into a plurality of partitions;

at step (b) storing keywords and document identifiers for the new document a predetermined partition of the supplemental portion;

at the predetermined intervals, sequentially merging the partitions of the supplemental portion with corresponding partitions of the main portion;

receiving bulk-load keywords and document identifiers for the index;

pre-dividing the bulk-load keywords and document identifiers into partitioned files related to the partitions of the main and supplemental portions of the index;

sequentially storing a partitioned file in the second storage device and merging the partition file with the corresponding partition of the main portion;

whereby bulk-load data may be efficiently integrated with the index.

70. The system of claim 57 wherein the partition file is merged with the corresponding partition of the main portion at a second predetermined interval different from the first predetermined interval.